Biodiversity Information Retrieval Through Large Scale Content-Based Identification: A Long-Term Evaluation

Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Jean-Christophe Lombardo, Robert Planqué, Simone Palazzo, Henning Müller

Biodiversity Information Retrieval Through Large Scale Content-Based Identification: A Long-Term Evaluation

Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Jean-Christophe Lombardo, Robert Planqué, Simone Palazzo and Henning Müller

Abstract Identifying and naming living plants or animals is usually impossible for the general public and often a difficult task for professionals and naturalists. Bridging this gap is a key challenge towards enabling effective biodiversity information retrieval systems. This taxonomic gap was actually already identified as one of the main ecological challenges to be solved during during the Rio de Janeiro United Nations "Earth Summit" in 1992. Since 2011, the LifeCLEF challenges conducted in the context of the CLEF evaluation forum have been boosting and evaluating the advances in this domain. Data collections with an unprecedented volume and diver-

Alexis Joly
Inria, LIRMM, Montpellier, France, e-mail: alexis.joly@inria.fr

Hervé Goëau
CIRAD, UMR AMAP, France, e-mail: herve.goeau@cirad.fr

Hervé Glotin
Universit de Toulon, Aix Marseille Univ, CNRS, LIS, DYNI team, Marseille, France, e-mail: herve.glotin@univ-tln.fr

Concetto Spampinato
University of Catania, Italy, e-mail: cspampin@dieei.unict.it

Pierre Bonnet
CIRAD, UMR AMAP, France, e-mail: pierre.bonnet@cirad.fr

Willem-Pier Vellinga
Xeno-canto foundation, The Netherlands, e-mail: wp@xeno-canto.org

Jean-Christophe Lombardo
Inria, LIRMM, Montpellier, France, e-mail: jean-christophe.lombardo@inria.fr

Robert Planqué
Xeno-canto foundation, The Netherlands e-mail: r.planque@vu.nl

Simone Palazzo
University of Catania, Italy e-mail: simone.palazzo@dieei.unict.it

Henning Müller
HES-SO, Sierre, Switzerland e-mail: henning.mueller@hevs.ch

sity have been shared with the scientific community to allow repeatable and long-term experiments. This paper describes the methodology of the conducted evaluation campaigns as well as providing a synthesis of the main results and lessons learned along the years.

1 Introduction

Identifying organisms is a key for accessing information related to the uses and ecology of species. This is an essential step in recording any specimen on earth to be used in ecological studies. Unfortunately, this is difficult to achieve due to the level of expertise necessary to correctly record and identify living organisms (for instance plants are one of the most difficult groups to identify with an estimated number of 400,000 species). This *taxonomic gap* has been recognized since the Rio Conference of 1992, as one of the major obstacles to the global implementation of the Convention on Biological Diversity. Among the diversity of methods used for species identification, Gaston and O'Neill (Gaston and O'Neill, 2004) discussed in 2004 the potential of automated approaches typically based on machine learning and multimedia data analysis. They suggested that, if the scientific community is able to (i) overcome the production of large training datasets, (ii) more precisely identify and evaluate the error rates, (iii) scale up automated approaches, and (iv) detect novel species, it will then be possible to initiate the development of a generic automated species identification system that could open up vistas of new opportunities for theoretical and applied work in biological and related fields. Since the question raised by Gaston and O'Neill (Gaston and O'Neill, 2004), *automated species identification: why not?*, a lot of work has been done on the topic (*e.g.* (Lee et al, 2004; Cai et al, 2007; Trifa et al, 2008; Towsey et al, 2012; Glotin et al, 2013a,b; Joly et al, 2014b)) and it is still attracting much research today, in particular on deep learning techniques. In parallel to the emergence of automated identification tools, large social networks dedicated to the production, sharing and identification of multimedia biodiversity records have increased in recent years. Some of the most active ones like eBird[1] (Sullivan et al, 2014), iNaturalist[2], iSpot (Silvertown et al, 2015), Xeno-Canto[3] or Tela Botanica[4], SABIOD and EADM CNRS[5] federations on machine learning for bioacoustics (respectively initiated in the US for the two first ones, and in Europe for the others), federate hundreds of thousands of active members, producing millions of observations each year. Noticeably, Pl@ntNet was the first initiative attempting to combine the force of social networks with automated identification tools (Joly et al, 2014b) through the release of a mobile application

[1] http://ebird.org/content/ebird/

[2] http://www.inaturalist.org/

[3] http://www.xeno-canto.org/

[4] http://www.tela-botanica.org/

[5] http://sabiod.org

and collaborative validation tools. As a proof of their increasing reliability, most of these networks have started to contribute to global initiatives on biodiversity, such as the Global Biodiversity Information Facility (GBIF[6]) which is the largest and most recognized one. Nevertheless, this explicitly shared and validated data is only the tip of the iceberg. The real potential lies in the automatic analysis of the millions of raw observations collected every year through a growing number of devices but for which there is no human validation at all. The performance of state-of-the-art multimedia analysis and machine learning techniques on such raw data (e.g., mobile search logs, soundscape audio recordings, wild life webcams, etc.) is still not well understood and is far from reaching the requirements of an accurate generic biodiversity monitoring system. Most existing research before LifeCLEF actually considered only a few dozen or up to hundreds of species, often acquired in well-controlled environments (Goëau et al, 2011a; Nilsback and Zisserman, 2008; Kumar et al, 2012). On the other hand, the total number of living species on earth is estimated to be around 10K for birds, 30K for fish, 400K for flowering plants (cf. State of the Worlds Plants 2017[7]) and more than 1.2M for invertebrates (Baillie et al, 2004). To bridge this gap, it is required to boost research on large-scale datasets and real-world scenarios. In order to evaluate the performance of automated identification technologies in a sustainable and repeatable way, the LifeCLEF[8] research platform was created in 2014 as a continuation of the plant identification task (Goëau et al, 2013b) that was run within the ImageCLEF lab [9] the three years before (Goëau et al, 2011a, 2012a, 2013a). LifeCLEF enlarged the evaluated challenge by considering birds and marine animals in addition to plants, and audio and video contents in addition to images. More concretely, the lab is organized around three tasks:

PlantCLEF: an image-based plant identification task making use of Pl@ntNet collaborative data, Encyclopedia of Life' data, and Web data

BirdCLEF: an audio recordings-based bird identification task making use of Xeno-canto collaborative data

SeaCLEF: a video and image-based identification task dedicated to sea organisms (making use of submarine videos and aerial pictures).

As described in more detail in the following sections, each task is based on big and real-world data and the measured challenges are defined in collaboration with biologists and environmental stakeholders so as to reflect realistic usage scenarios.

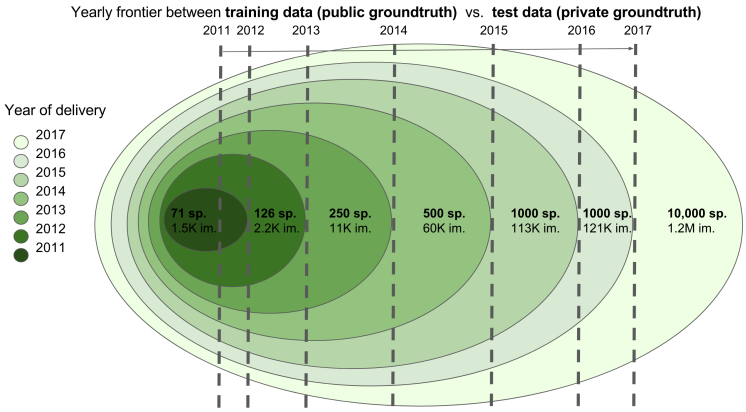

Fig. 1 Overview of the evaluation data used for the PlantCLEF challenge along the years

2 Plantclef: A 7-year-long Evaluation Of Image-based Plant Identification Systems

2.1 Methodology

The plant identification challenge of CLEF has been run since 2011, offering today a seven-year follow-up of the progress made in image-based plant identification. A particularity of the benchmark is that it always focused on real-world collaborative data contrary to many other test beds that were created beforehand in the context of well controlled laboratory conditions. Additionally, the evaluation protocol was defined in collaboration with biologists so as to reflect realistic usage scenarios. In particular, we considered the problem of classifying plant observations based on several images of the same individual plant rather than considering a classical image classification task. Indeed, it is usually required to observe several organs of a plant to identify it accurately (*e.g.* the flower, the leaf, the fruit, the stem, etc.). As a consequence, the same individual plant is often photographed several times by the same observer resulting in contextually similar pictures and/or near-duplicates. To avoid bias, it is crucial to consider such image sets as a single plant observation that should not be split across the training and the test set. In addition to the raw pictures, plant observations are usually associated with contextual and social data. This includes geo-tags or location names, time information, author names, collaborative ratings, vernacular names (common names), picture type tags, etc. Within all PlantCLEF challenges, the use of this additional information was considered as part

[6] http://www.gbif.org/

[7] https://stateoftheworldsplants.com/

[8] http://www.lifeclef.org

[9] http://www.imageclef.org/

of the problem because it was judged as potentially useful for a real-world usage scenario.

We provide in Fig.1 an overview of the data that was shared along the years within the PlantCLEF challenge. Each year, the data was considerably enriched and the number of species was increased from 71 species in 2011 to 10,000 species in 2017 (illustrated by more than 1 million images). This durable scaling-up was made possible thanks to the close collaboration of LifeCLEF with several important actors in the digital botany domain. First of all, the TelaBotanica social network. This network of expert and amateur botanists is one of the largest in the world (with about 40 thousand members) and is in charge of many citizen science projects relying on the collection of botanical observations by its members. TelaBotanica develops several collaborative tools dedicated to this purpose, in particular IdentiPlante [10] aimed at revising and validating the identification of the observations shared by the network. Most of the data used within the PlantCLEF challenge was collected and revised by the TelaBotanica network. Another source of data were contributions of the users of the Pl@ntNet application and the members of the TelaBotanica social network who validated many observations every year.

The evaluation metric that was used from 2011 to 2015 was an extension of the mean reciprocal rank (Voorhees et al, 1999), classically used in information retrieval. The difference is that it is based on a two-stage averaging rather than a flat averaging such as:

$$S = \frac{1}{U} \sum_{u=1}^{U} \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{r_{u,p}} \tag{1}$$

where U is the number of image authors within the test set, P_u the number of individual plants observed by the u-th author (within the test set), $r_{u,p}$ is the rank of the correct species within the ranked list of species returned by the evaluated system (for the p-th observation of the u-th author). If the correct species does not appear in the returned list, its rank $r_{u,p}$ is considered as infinite. Overall, the proposed metric makes it possible to compensate the long-tail distribution effects of social data. As in any social network, a few people actually produce huge quantities of data whereas the vast majority of contributors (the long tail) produce much less data.

2.2 Main Outcomes

Tables 1 and 2 give a year-to-year overview of the shared data and of the best performing systems (detailed descriptions of the results and systems can be found in the technical overview papers of each year (Goëau et al, 2011b, 2012b, 2013a, 2014, 2015, 2016, 2017) and participant working notes papers. To allow a comprehensive comparison along the years, we isolated in Table 1 the *leaf scans* and *white background* image categories that were part of the evaluation of the three first years but

[10] `http://www.tela-botanica.org/appli:identiplante` (in French)

that were abandoned afterwards. Table 2 focuses on photographs of plants in their natural environment (only leaves in 2011-2012, diverse organs and plant views in the following years). For a fair comparison, we also removed from the overview, the submissions that were humanly assisted in some point (*e.g.* involving a manual segmentation of the leaves).

Table 1 Three-year synthesis of the PlantCLEF challenge restricted to *leaf scans* and *pseudo-scans*.

Year	#Species	#Images	Evaluated systems	Score of best system	Brief description of best system
2011	71	3,967	20	0.574	Various local features (around Harris points) + Hash-based indexing + RANSAC based matching
2012	126	9,356	30	0.565	Shape and texture global features + SVM classifier
2013	250	11,031	33	0.607	Shape and texture global features + SVM classifier

Table 2 Seven-year synthesis of the results of the PlantCLEF challenge

Year	#Species	#Images	Evaluated systems	Performance of best system	Brief description of best system
2011	71	1,469	20	0.251	Model-driven segmentation Shape features . Random forest
2012	126	2,216	30	0.320	. Multi-scale local (color) texture SIFT + Sparse coding Spatial Pyramidal Matching . Linear SVM
2013	250	11,046	33	0.393	. Dense-SIFT, C-SIFT, Opponent SIFT HSV-SIF, self-similarity SSIM . Fisher Vectors . Linear Logistic Regression . Late fusion
2014	500	60,962	28	0.471	. ROI segmentation dense-SIFT + Color Moment . Fisher Vectors . SVM on FVs
2015	1000	113,205	18	0.667	. GoogLeNet CNN . 5-fold bagging + Borda fusion
2016	1000	121,205	29	0.827	. VGGNet, combine outputs of a same observation
2017	10,000	1,256,287	28	0.92	. Average of many fine-tuned CNNs

The main conclusion we can derive from the results of Table 1 is that the classical approach to plant identification consisting of analyzing the morphology of the leaves reached its limit. Leaf shape boundary features and shape matching techniques have been studied for 30 years and can be considered as sufficiently mature for capturing shape information in a robust and invariant way. The limited performance is thus rather due to the intrinsic limitation of using only the leaf morphology for discriminating a large number of species. The fact that scientists focused on leaf-based identification for many years is more related to the fact that the leaf was easier to scan and to process with state-of-the-art computer vision techniques of that period (segmentation, shape matching, etc.). With the arrival of more advanced computer vision techniques, we were progressively able to make use of other parts of the plant such as flowers or fruits, and to work on larger number of species. For this reason, metrics on leaf scans were abandoned from the PlantCLEF evaluation after 2013.

Table 2 gives the five-year synthesis of this approach to plant identification that we promoted through PlantCLEF. The most interesting conclusion we can derive is that we observed considerable improvements of the scores along the years whereas the difficulty of the task was increasing. The number of classes almost doubled every year between 2011 and 2015, starting from 71 species in 2011 and reaching 10,000 species in 2017. The increase of the performance can be explained by two major technological breakthroughs.

The first was the use of *aggregation-based* or *coding-based* image representation methods such as the Fisher Vector representation (Sánchez et al, 2013), which was used by the best performing system of Nakayama *et al.* (Nakayama, 2013) in 2013 and Qiang *et al.* (Chen et al, 2014) in 2014. These methods consist of producing high-dimensional representations of the images by aggregating previously extracted sets of hand-crafted local features into a global vector representation. They rely on a two step process: (i) the learning of a set of latent variables that explain the distribution of the local features in the training set (denoted as the codebook or vocabulary), and (ii) the encoding of the relationship between the local features of a given image and the latent variables. Overall, this allows to embed the fine-grained visual content of each image into a single representation space in which classes are easily separable even with linear classifiers.

The second technological step explaining the latest increase of performance is the use of deep learning methods, in particular convolutional neural networks (CNN) such as GoogLeNet (Szegedy et al, 2015). In 2015, the 10 best evaluated systems were based on CNNs. The performance difference is mainly due to particular system design improvements such as the use of bagging in the best run of Choi (Choi, 2015b). CNNs recently received a high amount of attention caused by the impressive performance they achieved in the ImageNet classification task (Krizhevsky et al, 2012). The force of these technologies relies on their ability to learn discriminant visual features directly from the raw pixels of the images without falling into the trap of the curse of dimensionality. This is achieved by stacking multiple *convolutional layers*, *i.e.* the core building blocks of a CNN. A convolutional layer basically takes images as input and produces as output *feature maps* corresponding to different convolution kernels, *i.e* looking for different visual patterns. Looking at the impressive

results achieved by CNN's in the 2015 edition of PlantCLEF there is absolutely no doubt that they are able to capture discriminant visual patterns of the plants in a much more effective way than previously engineered visual features. The editions of PlantCLEF 2016 and 2017 have also clearly confirmed the capacity of CNNs to take advantage of large noisy datasets. Indeed, in the 2017 edition, all networks trained solely on the noisy dataset (coming from web crawl) outperformed the same models trained on the trusted data (coming from the trusted Encyclopedia of Life website). Even at a constant number of training iterations (*i.e.* at a constant number of images passed to the network), it was more profitable to use the noisy training data. This means that diversity in the training data is a key factor to improve the generalization ability of deep learning. The noise itself seems to act as a regularization of the model. The amazing performance of the best runs, which reached a score higher than 90% of correct identification on 10,000 classes opens new perspectives on the potential of automated plant species capacities at the world level.

3 Birdclef: A 4 Year Long Evaluation Of Bird Sound Identification Systems

3.1 Methodology

The bird identification challenge of LifeCLEF, initiated in 2014 in collaboration with Xeno-Canto, considerably increased the scale of the seminal challenges. The first bird challenge ICML4B (Glotin et al, 2013a) initiated in 2012 by DYNI/SABIOD had only 35 species, but received 400 runs. The next at MLSP had only 15 species, the third (NIPS4B(Glotin et al, 2013b) in 2013 by SABIOD) had 80 species. Meanwhile, Xeno-canto, launched in 2005, hosts bird sounds from all continents and daily receives new recordings from some of the remotest places on Earth. It currently archives with 379472 recordings, 9779 species of birds, making it one of the most comprehensive collections of bird sound recordings worldwide, and certainly the most comprehensive collection shared under Creative Commons licenses.

For the BirdCLEF challenge, it was decided to not consider the whole Xeno-Canto dataset but to rather focus on a specific region. The objective was to find a good trade-off between scalability and biodiversity coverage. A sufficient number of species had to be considered so as to evaluate the feasibility of a real-world biodiversity monitoring system. But on the other side, it was necessary to limit the volume of data to be processed by the participating research groups so as to mitigate computational challenges and data management. The chosen region of interest has been the Amazonian rain forest because it is one of the richest in the world in terms of biodiversity but also one of the most endangered. For the first edition of the challenge, in 2014, the evaluation dataset was restricted to the 500 species having the most records in an Amazonian area straddling Brazil and neighboring countries. The geographical extent and the number of species were progressively increased over the

years so as to reach 1000 species in 2015/2016, and 1500 in 2017. By nature, the Xeno-Canto data as well as the BirdCLEF subset has a massive class imbalance. For instance, the 2017 dataset contains 48,843 recordings in total, with a minimum of four recordings for *Laniocera rufescens* and a maximum of 160 recordings for *Henicorhina leucophrys*.

A comprehensive overview of the data shared[11] over the years is provided in Fig.2. Each year, selected Xeno-canto recordings were split in two parts: 2/3 of the data was shared as training data so as to allow participants to train and optimize their system, and the other 1/3 of the recordings were kept as official test samples and shared to the participants a few weeks after the training set. To avoid participants tuning their system on the test data, the species names were removed from the test set and kept secret over the years (*i.e.* participants have to run their system in a blind manner). To allow a long-term evaluation of the progress made, it was also ensured that the test data provided each year were a superset of the test data of the previous years. Furthermore, the recordings were shared using a stable format along the years. Each audio file was associated with an XML file containing the available meta-data such as the date, the geo-location, the author, the type of sound (call, song, alarm, flight, etc.) or some collaborative quality ratings. For the training set, the meta-data also included the information related to the species of the bird(s) vocalizing within the recording (taxonomic names and sometimes common names). Most Xeno-Canto recordings are captured using mono-directional devices in order to focus on a single vocalizing bird. The name of the species of this primary singing bird is annotated in the meta-data through a field entitled "foreground species". But often, there is also a number of other birds that can be heard in the background. The names of the species of the other birds are often annotated in the meta-data through a field entitled "background species".

Identifying birds from mono-directional recordings such as the ones discussed above is of high interest for many scenarios. In particular, this could help non-experts as well as experts in the process of collecting and identifying such new recordings. To complement this, there is also an interest in identifying birds from *omnidirectional* recordings (i.e. the target is the foreground species) or *soundscape*. This enables more passive monitoring scenarios such as setting up a network of static recorders that would continuously capture the surrounding sound environment. Therefore, we started to integrate soundscape recordings within the BirdCLEF challenge in 2016. A significant number of recordings tagged as *soundscapes* actually already existed in the Xeno-Canto collection. They usually correspond to longer recordings than the mono-directional ones and they do not have any *foreground* species in the meta-data. 925 of such soundscapes were found in the Amazonian area and were integrated as a new test within the BirdCLEF 2016 challenge. One of the limitations of this new content, however, was that the vocalizing birds were not localized in the recordings. The set of species audible in the recording was identified in the meta-data but the vocalizing specimens were not localized in time. Thus, to allow a more accurate evaluation, it was decided to introduce new time-coded

[11] some sample can be listen at `http://sabiod.org/DYNITAG/BIRDCLEF`

soundscapes within the BirdCLEF 2017 challenge. In total, 6.5 hours of recordings were collected in the Amazonian forests and were manually annotated by two experts including a native of the Amazon forest, in the form of time-coded segments with associated species name.

The evaluation protocol of BirdCLEF remained roughly the same during the 4 years it ran. Participants were asked to run their system so as to identify all the actively vocalizing bird species in each test recording (or in each test segment of 5 seconds for the soundscape). Up to 4 *run files* per participant could be submitted to allow evaluating different systems or system configurations (a *run file* is a formatted text file containing the species predictions for all test items). Each species had to be associated with a normalized score in the range $[0,1]$ reflecting the likelihood that this species is singing in the test sample. For each submitted run, participants had to signal if the run was performed fully automatically or with human assistance, and if they used a method based only on audio analysis or with the use of the metadata. The evaluation metric used was the mean Average Precision (mAP) averaged across all queries, considering each recording in the test set as a query and computed as:

$$mAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q},$$

where Q is the number of test samples and $AveP(q)$ for a given test file q is computed as

$$AveP(q) = \frac{\sum_{k=1}^{n} (P(k) \times rel(k))}{\text{number of relevant documents}}.$$

Here k is the rank in the sequence of returned species, n is the total number of returned species, $P(k)$ is the precision at cut-off k in the list and $rel(k)$ is an indicator function equaling 1 if the item at rank k is a relevant species (i.e. one of the species in the ground truth).

3.2 Main Outcomes

Between 60 and 90 research groups registered each year for the BirdCLEF challenge and about 20 of them submitted run files to at least one of the yearly campaigns (with a variation of 5 to 10 participants depending on the year). This durable evaluation allowed to accelerate the progress made along the years by measuring it accurately thanks to the re-used sub-set test data. As a synthesis of this long-term effort, Fig. 3 displays the evolution of the best mAP scores that were obtained over the years. The curve corresponding to the 2014 test set, in particular, shows the impressive progress that was made from the beginning of the challenge. The best mAP value actually increased from 0.51 to 0.76 in 4 years (for the mono-directional recordings). A big step was particularly observed between 2015 and 2016 (Goëau et al, 2016). It was exclusively due to the progress of the underlying methods and algorithms since the training set shared within BirdCLEF was exactly the same for these

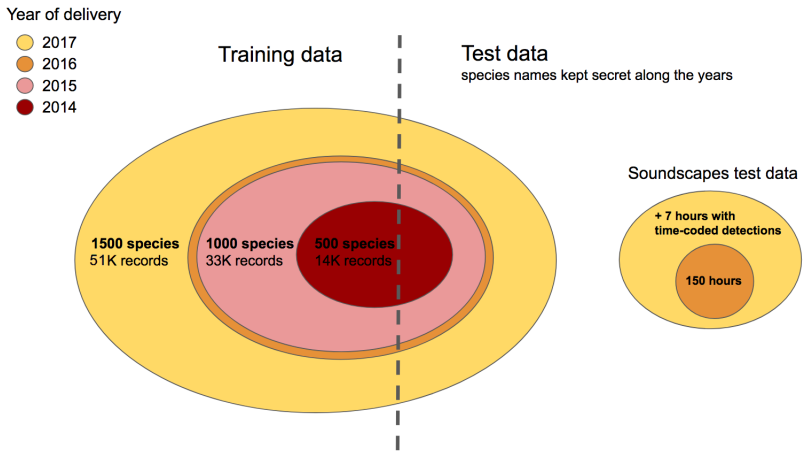

Fig. 2 Overview of the evaluation data used for the BirdCLEF challenge along the years

two years (as illustrated in Fig.2). More precisely, and without great surprise, the best system evaluated in 2016 was the first one using deep learning technologies. The convolutional neural network it relied on, outperformed by 4% the mAP of the previous state-of-the-art method of 2014 and 2015, which was based on strong feature engineering and classical machine learning algorithms. After this first remarkable success, most participants in the BirdCLEF challenge continued exploring the use of CNNs in 2017. The different systems used in 2017 mainly differed in the employed CNN architecture and in the time-frequency representation given as input of the CNN. Interestingly, the best system in 2017, from DYNI LSIS CNRS team (Sevilla and Glotin, 2017; Joly et al, 2017), was an adaptation of the Inception model (version 4), *i.e.* a CNN that was designed by Google for large scale image classification tasks. This raw model was fine-tuned directly from the weights of the initial image classifier. This illustrates the strong convergence of machine learning methods for different contents and the feasibility of transferring knowledge from one modality to another, as long as one uses a common representation (i.e. 2D time-frequency images). The second main outcome of BirdCLEF which can be observed in Fig. 3, is that the soundscape task appears to be much more challenging than the classical task that we shall consider here as mono-species recordings. The identification performance actually remains pretty high for the mono-species recordings, even when considering all the species vocalizing in the background (yellow curve). On the contrary, the best mAP obtained on the 2016 soundscape data set is very low and did not improve between 2016 and 2017 (red curve). One of the main difficulties of such recordings is that many individual birds of several species are often singing simultaneously. This profusion of overlapping sources causes the classical CNN models trained on the mono-species to fail. A good method on the soundscape task seemed to be the feature engineering based method of Lasseck et al. (Lasseck, 2015), as the deep learning methods employed by the other participants in 2016

and 2017 were less efficient on the non time-coded soundscape 2016 test set. It is likely that the strength of the features engineering method is based on the extraction of very species-specific time-frequency features. This expert fine-grained approach may allow the extraction of features more robust to the species overlap problem. This verdict was the main reason why we introduced a new soundscape dataset in 2017 (Goëau et al, 2017), in the form of time-coded segments of 5 seconds, each associated with the list of species vocalizing in this small segment. The goal was to encourage the participants to output predictions at that temporal resolution instead of processing the whole soundscape as a classical recording. The performance achieved on this new test set (dark red point Fig.3) confirmed that the temporal resolution of the prediction was one of the issues and that processing each chunk of 5 seconds separately improves the results over the previous soundscape test set. However, the best performance remains much lower than for the mono-species task. One of the most likely reasons is the bias between the training data (mono-species) and the test data (soundscape). The overlap of all the birds vocalizing simultaneously actually induces audio patterns that cannot be captured directly from the mono-species recordings. A solution to learn such patterns would be to integrate soundscape with time-coded annotations in the training set itself. This approach is unfortunately not realistic because of the cost to produce such content. Another more realistic perspective is to run data augmentation synthesizing new training data from the mono-species recordings themselves. The improvement of the quality of automatic bird activity detection (BAD) is also being taken in consideration as recently depicted in the BAD challenge (Stowell et al, 2016). Finally, we are investigating a more advanced paradigm towards binaural source diarization and joint classification from stereo soundscape in future BirdCLEF sessions.

4 SeaCLEF: A 4-year Evaluation Of Sea Organisms Identification

The need for automated methods for sea-related multimedia data is driven by the recent sprout of marine and ocean observation approaches (mainly imaging systems) and their employment for marine ecosystem analysis and biodiversity monitoring. Indeed in recent years we have witnessed an exponential growth of sea-related multimedia data in the forms of images/videos/sounds, for disparate reasons ranging from fish biodiversity monitoring to marine resource managements to fishery to educational purposes. However, the analysis of such data is particularly expensive for human operators, thus limiting the impact that the technology may have in understanding and sustainably exploiting the sea/ocean. Within LifeCLEF, we investigated several highly demanding annotation scenarios including coral reef fish species monitoring, humpback whale individual recognition, salmon detection for water turbine monitoring and picture-based marine animal species recognition. In the following two subsections, we give an overview of the two challenges that

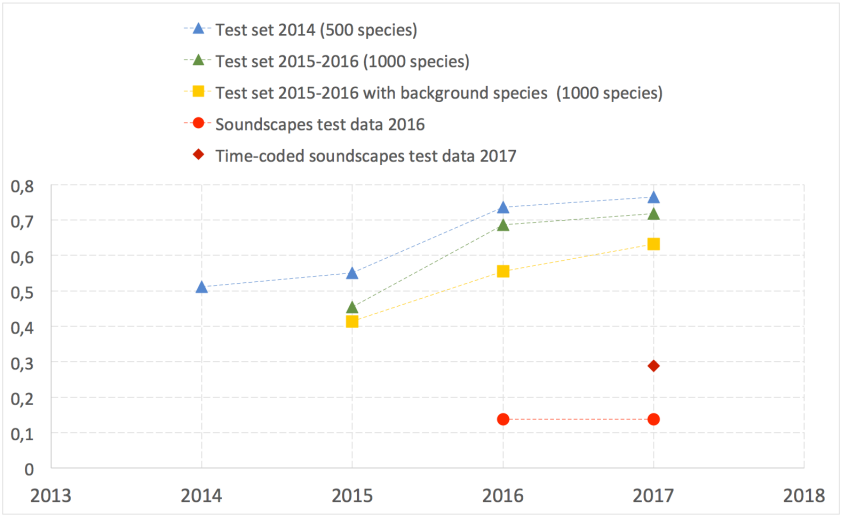

Fig. 3 Overview of the performance of the best systems evaluated within the BirdCLEF challenge (for different test data sets)

attracted the most participants and that were conducted over several consecutive years.

4.1 Underwater Coral Reef Species Monitoring: Methodology And Main Outcomes

Underwater imaging systems are increasingly used in a range of monitoring or exploratory applications, in particular for biological (e.g. benthic community structure, habitat classification), fisheries (e.g. stock assessment, species richness), geological (e.g. seabed type, mineral deposits) and physical surveys (e.g. pipelines, cables, oil industry infrastructure). Their usage has benefitted from the increasing miniaturization and cost-effectiveness of submersible ROVs (remotely operated vehicles) and advances in underwater digital cameras. These technologies have revolutionized our ability to capture high-resolution images in challenging aquatic environments and are also greatly improving our ability to effectively manage natural resources, increasing our competitiveness and reducing operational risks in industries that operate in both marine and freshwater systems. Despite these advances, the analysis of the produced data usually requires very time-consuming and expensive input by human observers. This is particularly true for ecological and fishery video data, which often requires laborious visual analysis. This analytic bottleneck greatly restricts the use of these otherwise powerful video technologies and demands effective methods for automatic content analysis to enable proactive provision of analytic informa-

tion. The underwater video dataset used within LifeCLEF was derived from the Fish4Knowledge video repository, which contains about 700,000 10-minute video clips that were taken in the past five years to monitor Taiwan's coral reefs. The Taiwan area is particularly interesting for studying the marine ecosystem, as it holds one of the largest fish biodiversities of the world with more than 3,000 different fish species [12]. The dataset contains videos recorded from sunrise to sunset showing several phenomena, e.g. murky water, algae on camera lens, etc., which make the identification task more complex. Each video has a resolution of either 320x240 or 640x480 with 5 to 8 fps.

The data set used for the coral reef challenge of LifeCLEF 2015, LifeCLEF 2016 and LifeCLEF 2017 was a small annotated subset of the Fish4Knowledge repository (Spampinato et al, 2016). It was composed of about 90 videos manually annotated for a list of 15 fish species. Each video was labelled and agreed by two expert annotators and the ground truth consists of a set of bounding boxes (one for each instance of the given fish species list) together with the fish species. In total the dataset contained more than 9000 annotations (bounding boxes + species) with a relatively high imbalance in the number of instances of fish species: for instance it contained 3165 instances of "Dascyllus Reticulates" and only 72 instances of "Zebrasoma Scopas". For each considered fish species, its fishbase.org link was also given. In the fishbase webpage, participants could find more detailed information about fish species including also high quality images that could be used as additional training data. In order to make the identification process independent from tracking, temporal information was not exploited. This means that the annotators only labelled fish for which the species was clearly identifiable, i.e., if at frame t the species of fish A is not clear, it was not labelled, no matter if the same fish was in the previous frame (t-1). Each video was accompanied by an xml file that contains instances of the provided list species as well as information on the camera location *e.g.*

```
<?xml version="1.0" encoding="utf-8"?>
<video id="0b21f0579d247c855e05405d3ed805c1#201205251240" location="NPP3" camera="4">
   <frame id="0">
      <object fish_species="Dascyllus Aruanus" h="68" w="87" x="322" y="233"/>
   </frame>
   <frame id="1">
      <object fish_species="Dascyllus Aruanus" h="68" w="87" x="319" y="230"/>
   </frame>
   <frame id="2">
      <object fish_species="Dascyllus Aruanus" h="68" w="87" x="342" y="231"/>
   </frame>
   <frame id="391">
      <object fish_species="Plectrogly-Phidodon Dickii" h="50" w="35" x="271" y="336"/>
      <object fish_species="Plectrogly-Phidodon Dickii" h="41" w="29" x="339" y="375"/>
   </frame>
</video>
```

Since the end-to-end objective of the task was to count the number of specimens per species (for biodiversity monitoring), we introduced two related evaluation metrics: the "**Counting Score (CS)**" and the "**Normalized Counting Score (NCS)**",

[12] for which a taxonomy is available at `http://fishdb.sinica.edu.tw`

defined as:

$$CS = e^{-\frac{d}{N_{gt}}} \tag{2}$$

with d being the difference between the number of occurrences in the run (per species) and, N_{gt}, the number of occurrences in the ground truth. The Normalized Counting Score instead depends on precision Pr:

$$NCS = CS \times Pr = CS \times \frac{TP}{TP+FP} \tag{3}$$

with $Pr = TP/(TP+FP)$, TP and FP being the true positives and the false positives. A detection was considered as true positive if the intersection over union score of its bounding box and the ground truth was over 0.5 and the species was correctly identified.

Fig. 4 shows an overview of the score obtained by the best systems evaluated in 2015, 2016 and 2017 on the same coral reef data set. In the previous fish classification challenge the hierarchical LBP classifier (DYNI team (Joalland et al, 2014)) won. However, CNN was the best system of 2015 (by SNUMED (Choi, 2015a)), and it was not outperformed in the following years. Contrary to all other LifeCLEF challenges, no real progress were thus observed over the years. The system of SIATMMLAB in 2017 (Zhuang et al, 2017) was devised as an improvement of the one of SNUMED but its precision was still lower resulting in a lower Normalized Counting Score in the end. The right plots of Fig.4 show that the main strength of the SNUMED system is to be more stable than the other systems across the different species. Importantly, this is rather due to a better detection of the candidate fish instances than a better performance of the classification of the resulting bounding boxes. The SIATMMLAB system actually used a more advanced convolutional neural network model for the classification but it was less accurate in the preliminary detection phase.

4.2 Individual Whale Identification: Methodology And Main Outcomes

The problem of automatically identifying individual organisms rather than species has received much less attention. Yet, for some groups, it is preferable to monitor the organisms at the individual level rather than at the species level. This is notably the case of big animals, such as whales and elephants, of which the populations are scarcer and are traveling longer distances. Monitoring individual animals allows gathering valuable information about population sizes, migration, health, sexual maturity and behavior patterns. Tracking devices and tagging technologies are only part of the solution because of their invasive character, relatively high cost and limited lifetime. Morphological/biometric approaches are a complementary approach

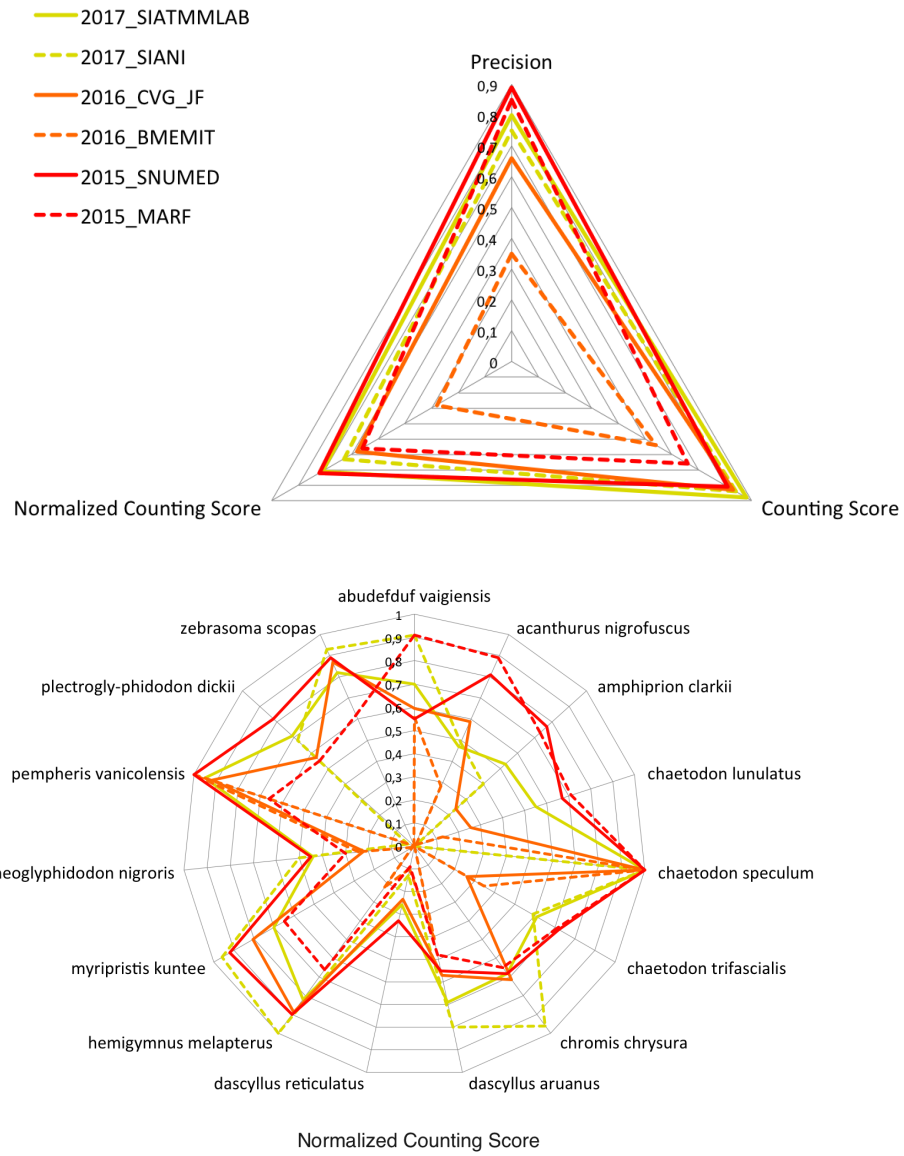

Fig. 4 Overview of the performance of the best systems evaluated within the coral reef species recognition challenge

that is less invasive, more durable and cheaper for nature watchers mobilized on a given spot. Using natural markings to identify individual animals over time is usually known as photo-identification. This research technique is used on many species

of marine mammals. Initially, scientists used artificial tags to identify individual whales, but with limited success (most tagged whales were actually lost or died). In the 1970s, scientists discovered that individuals of many species could be recognized by their natural markings. These scientists began taking photographs of individual animals and comparing these photos against each other to identify individual animal movements and behavior over time. Since its development, photo-identification has proven to be a useful tool for learning about many marine mammal species including humpbacks, right whales, finbacks, killer whales, sperm whales, bottlenose dolphins and other species to a lesser degree. This process is still mostly done manually making it impossible to get an accurate count of all the individuals in a given large collection of observations. Researchers usually survey a portion of the population, and then use statistical formulae to determine population estimates. To limit the variance and bias of such an estimator, it is however required to use sufficiently large samples that still make it a very time-consuming process. Automating the photo-identification process could drastically scale-up such surveys and open brave new research opportunities for the future.

To evaluate this scenario, we did set up a test-bed in collaboration with Cetamada, a Malagasy Non-Profit Association created in May 2009, whose goal is to protect marine mammal population and their habitat in Madagascar through sustainable ecotourism and scientific research. There are presently 4 citizen sciences data collection sites (St. Marys, Majunga, Ifaty and Fort Dauphin) for which hotel-establishments and their customers have become sentinels for data collection. This method helps obtain more than 250 photo IDs each year, which effectively helps produce a photo catalogue of humpback whales reproducing on Malagasy coasts. From that data, we built an evaluation dataset of 2005 images of humpback whales that were collected between 2009 and 2014. After acquisition, each photograph was manually cropped so as to focus only on the caudal fin that is the most discriminant pattern for distinguishing one individual whale from another. Actually, the fins can be distinguished thanks to the natural markings and/or the scars that appear along the years. Automatically finding such matches in the whole dataset and rejecting the false alarms is difficult for three main reasons. The first reason is that the number of individuals in the dataset is high, around $1,200$, so that the proportion of true matches is actually very low (around 0.05% of the total number of potential matches). The second difficulty is that distinct individuals can be very similar at a first glance and that it is often difficult to distinguish them even for a human annotator. To discriminate the true matches from such false positives, it is required to detect very small and fine-grained visual variations such as in a spot-the-difference game. The third difficulty is that all images have a similar water background of which the texture generates quantities of local mismatches.

Concretely, the task consisted in detecting as many true matches as possible from the whole dataset, in a fully unsupervised way. Each evaluated system had to return a *run file* (i.e., a raw text file) containing as many lines as the number of discovered matches, each match being a triplet of the form:

$$< imageX.jpg\ imageY.jpg\ score >$$

where *score* is a confidence score in $[0,1]$ (1 for highly confident matches). The retrieved matches had to be sorted by decreasing confidence score. A run should not contain any duplicate match (e.g., $<$ *image*1.*jpg image*2.*jpg score* $>$ and $<$ *image*2.*jpg image*1.*jpg score* $>$ should not appear in the same run file). The metric used to evaluate each run was Average Precision:

$$AveP = \frac{\sum_{k=1}^{K} P(k) \times rel(k)}{M}$$

where M is the total number of true matches in the groundtruth, k is the rank in the sequence of returned matches, K is the number of retrieved matches, $P(k)$ is the precision at cut-off k in the list, and $rel(k)$ is an indicator function equaling 1 if the match at rank k is a relevant match, 0 otherwise.
The same challenge was run for two consecutive years, in 2016 and 2017. An overview of the results achieved by the best system of each participant (yearly) is provided in Fig.5. In 2016, the best result was achieved by the INRIA-ZENITH team who used a large-scale matching system based on SIFT visual features, approximate k-nn search and a RANSAC-like spatial consistency refinement step to reject false positives (Joly et al, 2016). In 2017, a similar system was re-implemented by BMEMIT and extended with an additional clustering step which provided a consistent improvement (Dvid Papp and Szcs, 2017). Interestingly, the whale photo-identification challenge is the only one within LifeCLEF for which deep learning technologies do not provide the best performance (although several attempts were made). The main reason is that it is very different from the classical challenges studied in the machine learning community. This is actually an unsupervised classification problem but for which the visual patterns to be discovered are very small and lost among a high amount of other highly similar patterns. Only an explicit spatial verification based on the hypothesis of an epipolar geometry allows to distinguish the real matches from the distractors. Without supervision, convolutional neural networks fail to capture this property.

5 Cross-task Analysis Of The Use Of Contextual Meta-data

Most of the data sets shared within LifeCLEF since 2011 included contextual meta-data in addition to the raw audio-visual contents. As an illustration, Tab.3 lists the meta-data shared for each image of the training set of PlantCLEF 2016. A large fraction of the plant and bird observations, in particular, were associated with their date and geo-location. This information was expected to be highly useful for species identification. Indeed, most plants and animals live in specific ecological niches and are likely to be observed at some specific periods.

Table 4 reports the results obtained by the participants of the plant and the bird tasks who attempted to evaluate the potential benefit of this meta-data over the years. However, the benefit of using the temporal and spatial information has never been

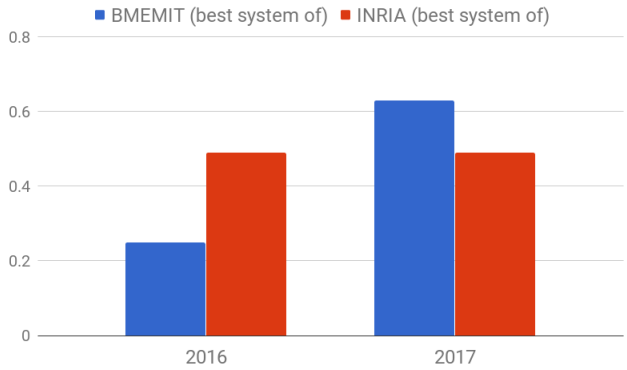

Fig. 5 Overview of the performance of the best systems evaluated within the whale photo-identification challenge

Table 3 Types of metadata shared within PlantCLEF challenge

Type of metadata	Metadata description
ObservationId	the plant observation ID from which several pictures can be associated
MediaId	the ID of the image
View Content	Description of the content visible in the image : Entire or Branch or Flower or Fruit or Leaf or LeafScan, etc.
ClassId	the class number ID that must be used as ground-truth. It is a numerical taxonomical number used by Tela Botanica
Species name	the species names (containing 3 parts: the Genus name, the specific epithet, the author(s) who discovered or revised the name of the species)
Family	the name of the Family, two levels above the Species in the taxonomical hierarchy used by Tela Botanica
Date	(if available) the date when the plant was observed
Vote	the (round up) average of the user ratings of image quality
Location	(if available) locality name, a town most of the time
Latitude & Longitude	(if available) the GPS coordinates of the observation in the EXIF metadata, or if no GPS information were found in the EXIF the GPS coordinates of the locality where the plant was observed (only for the towns of metropolitan France)
Author	name of the author of the picture
YearInCLEF	ImageCLEF2011, ImageCLEF2012, ImageCLEF2013, PlantCLEF2014, PlantCLEF2015

decisive in any of the LifeCLEF challenges. Worse, it often degraded the performance compared to using the raw audio-visual data solely. To better highlight this finding, Table 4 provides an overview of all the experiments for which it was possible to evaluate the performance of the same system with or without the use of meta-data. The best improvement was achieved by the Inria team in 2013 for the plant task and 2014 for the bird task. Both were obtained by post-filtering the list of

candidate species based on a temporal histogram constructed for each species based on the training meta-data. However, these runs were still outperformed by purely content-based methods developed by other participants.
This difficulty of successfully using geography and seasonality is quite surprising. It is actually accepted that the habitat of a given species is highly correlated with its ecological profile. Several reasons explain this paradox. The first one is that the occurrence data of the training set is too sparse to accurately model the distribution of the species. The second reason is that the used machine learning techniques were too straightforward to well address the problem. As discussed in Section 6, species distribution modeling from occurrence data is still a hard problem in ecology, in particular in the context of uncontrolled observations such as the one used in the PlantCLEF challenge.
Concerning the use the observation date, which was the second most used meta-data by participants, there is several difficulties to appropriately exploit it. First, the plant phenology (plant life cycle events) for a given species is different according to its location (*i.e* the same species will present different flowering periods, if individuals are not at the same altitude in mountain conditions, are not exposed along the year to the same light conditions, etc.). Secondly, it's now well accepted that the plant phenology for a given species is changing from one year to another one, according to the climate changes. It is then difficult to find a regular pattern over several years, even if observations are produced at the same location. Thirdly, as plant phenology is profoundly influenced by human activity (fertilizer, pruning, greenhouse cultivation, etc.), the phenology of most of the plants observed in urban areas can be different than the individuals growing in natural conditions. According to these various factors, and the limited number of observations per species, one can understand that it is not easy to find a method which is robust on a large scale for a strong improvement of the identification performance. The potential of the use of meta-data, which is recognized as highly relevant by naturalists, has still to be demonstrated, and will be a central part of a new challenge entitled *GeoLifeClef*, that will be launched in 2018.

Table 4 Impact of the use of metadata for plants and birds identification

Year	Task	Team	Metadata type	Improvement
2011	PlantCLEF	UAIC	GPS, Date, Author Id	-35.89%
2012	PlantCLEF	BTU DBIS (Böttcher et al, 2012)	GPS	-4.76%
2013	PlantCLEF	Inria (Bakic et al, 2013)	Date	**+9.06%**
2015	PlantCLEF	SABANCI-OKAN (Ghazi and Ozdemir, 2015)	Date	**+1.23%**
2014	BirdCLEF	Inria (Joly et al, 2014a)	GPS, Date	**+11.28%**
2017	BirdCLEF	TUCMI (Kahl et al, 2017)	GPS	-32.67%

6 GeoLifeClef: A Machine Learning Approach to species Distribution Modeling

In order to increase the interest of the computer science community in the use of the *undisclosed* potential of meta-data for automated species identification, we designed a new challenge within LifeCLEF to be ran in 2018 for the first time. In particular, the new task called *GeoLifeClef* will focus on *location-based species recommendation*. Automatically predicting the list of species that are the most likely to be observed at a given location is useful for many scenarios in biodiversity informatics: a) it could improve species identification processes and tools by reducing the list of candidate species that are observable at a given location; b) it could facilitate biodiversity inventories through the development of location-based recommendation services (typically on mobile phones) as well as the involvement of non-expert nature observers; c) last but not least, it might serve educational purposes thanks to biodiversity discovery applications providing functionalities such as contextualized educational pathways. This new challenge will contribute to increase exchanges between the computer science community and ecological statisticians working on species distribution modelling problems, who would both have lots to gain by sharing their experiences and knowledge.
Concretely, the objective of the challenge will be to predict the list of species that are the most likely to be observed at a given location. Therefore a large training set of species occurrences will be provided, each occurrence being associated with a multi-channel image characterizing the local environment. Indeed, it is usually difficult to learn a species distribution model directly from spatial positions because of the limited number of occurrences and the sampling bias. What is usually done in ecology is to predict the distribution on the basis of a representation in the environmental space, typically a feature vector composed of climatic variables (average temperature at that location, precipitation, etc.) and other variables such as soil type, land cover, distance to water, etc. As illustrated in Fig. 6 the originality of GeoLifeCLEF is to generalize such a niche modeling approach to the use of an image-based environmental representation space. Instead of learning a model from environmental feature vectors, the goal of the task will be to learn a model from k-dimensional image patches, each patch representing the value of an environmental variable in the neighborhood of the occurrence. From a machine learning point of view, the challenge will thus be treatable as an image classification task.

According to the huge volume of new data produced by large scale citizen science initiatives, such as eBird, iNaturalist, or Pl@ntNet, and the accessibility of various environmental data based on the open science movement, the adaptation potential (to various living organism groups, environments, regions, etc.) of the result of this task is extremely important. The hope with this new task is to open new interdisciplinary research opportunities based on the analysis of a very large amount of data that was never mobilized beforehand.

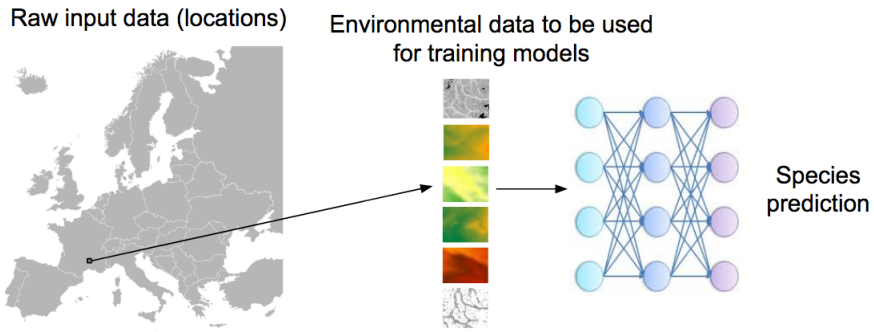

Fig. 6 Overview of the GeoLifeClef challenge

7 Conclusion

This chapter has discussed the experience of running the LifeCLEF challenges from 2011 to 2017. Several large-scale and repeatable experiments were designed over the years in order to boost research on biodiversity information retrieval. A high number of research groups participated in and benefited from this joint research effort. Overall, LifeCLEF has had an important impact in different fields including multimedia information retrieval, machine learning and biodiversity informatics (more than 500 citations at the end of 2017 according to Google scholar). The main lessons we learned in the design of attractive, sustainable and impacting challenges are the following:

- **Data is a key factor:** sharing original, valuable and large-scale data sets is a key factor for attracting researchers on a given challenge. Within LifeCLEF, tens of men months have been spent in integrating, cleaning and annotating the raw content of data providers.
- **Hard problems but simple tasks**: if the task is too specific or too complex in terms of objectives, it is not attractive. For instance, it is crucial to avoid fragmenting the challenge in many subtasks even if at a first glance it can appear as a good way to better understand the results. What happens in practice is that the participation is fragmented as well: only a few systems are run for each subtask and there is not enough output data to conduct relevant analyses. A single task relying on a hard scientific problem is the best way to federate a community around a given topic.
- **Sustaining the community requires a good trade-off between novelty and continuity:** research relies on long-term efforts and investigations. Thus, it is important to avoid switching to a new problem when the previous one is not solved. On the other hand, sticking exactly to the same challenge over years is counterproductive in terms of attractiveness and emulation. The good trade-off consists in progressively increasing the complexity and/or the difficulty of the

task but preserving a sufficient continuity to allow former participants to build on top of their acquired knowledge and technologies.

Acknowledgements: The organization of the PlantCLEF task is supported by the French project FlorisTic (Tela Botanica, INRIA, CIRAD, INRA, IRD) funded in the context of the national investment program PIA. The organization of the BirdCLEF task is supported by the Xeno-Canto foundation for nature sounds as well as the French CNRS project SABIOD.ORG and EADM MADICS, and FlorisTic. The annotations of some soundscapes were prepared with the late wonderful Lucio Pando at Explorama Lodges, with the support of Pam Bucur, Marie Trone and H. Glotin. The organization of the SeaCLEF task is supported by the Ceta-mada NGO and the French project FlorisTic.

References

Baillie J, Hilton-Taylor C, Stuart SN (2004) 2004 IUCN red list of threatened species: a global species assessment. Iucn

Bakic V, Mouine S, Ouertani-Litayem S, Verroust-Blondet A, Yahiaoui I, Goëau H, Joly A (2013) Inria's participation at imageclef 2013 plant identification task. In: CLEF (Online Working Notes/Labs/Workshop) 2013

Böttcher T, Schmidt C, Zellhöfer D, Schmitt I (2012) Btu dbis'plant identification runs at imageclef 2012. In: CLEF (Online Working Notes/Labs/Workshop)

Cai J, Ee D, Pham B, Roe P, Zhang J (2007) Sensor network for the monitoring of ecosystem: Bird species recognition. In: Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on, DOI 10.1109/ISSNIP.2007.4496859

Chen Q, Abedini M, Garnavi R, Liang X (2014) Ibm research australia at lifeclef2014: Plant identification task. In: Working notes of CLEF 2014 conference

Choi S (2015a) Fish identification in underwater video with deep convolutional neural network: Snumedinfo at lifeclef fish task 2015. In: CLEF (Working Notes)

Choi S (2015b) Plant identification with deep convolutional neural network: Snumedinfo at lifeclef plant identification task 2015. In: Working notes of CLEF 2015 conference

Dvid Papp FM, Szcs G (2017) Image matching for individual recognition with sift, ransac and mcl. In: Working Notes of CLEF 2017 (Cross Language Evaluation Forum)

Gaston KJ, O'Neill MA (2004) Automated species identification: why not? Philosophical Transactions of the Royal Society of London B: Biological Sciences 359(1444):655–667

Ghazi EAOM Berrin Yanikoglu, Ozdemir MC (2015) Sabanci-okan system in lifeclef 2015 plant identification competition. In: Working notes of CLEF 2015 conference

Glotin H, Clark C, LeCun Y, Dugan P, Halkias X, Sueur J (2013a) Proc. 1st workshop on Machine Learning for Bioacoustics - ICML4B. ICML, Atlanta USA, URL http://sabiod.org/ICML4B2013_book.pdf

Glotin H, LeCun Y, Artières T, Mallat S, Tchernichovski HX O (2013b) Proc. Neural Information Processing Scaled for Bioacoustics, from Neurons to Big Data. NIPS Int. Conf., Tahoe USA, URL http://sabiod.org/nips4b

Goëau H, Bonnet P, Joly A, Boujemaa N, Barthélémy D, Molino JF, Birnbaum P, Mouysset E, Picard M (2011a) The imageclef 2011 plant images classification task. In: CLEF 2011

Goëau H, Bonnet P, Joly A, Boujemaa N, Barthelemy D, Molino JF, Birnbaum P, Mouysset E, Picard M (2011b) The CLEF 2011 Plant Images Classification Task. In: Petras V, Forner P, Clough P, Ferro N (eds) CLEF 2011 Working Notes, CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-1177/

Goëau H, Bonnet P, Joly A, Yahiaoui I, Barthélémy D, Boujemaa N, Molino JF (2012a) Imageclef2012 plant images identification task. In: CLEF 2012, Rome

Goëau H, Bonnet P, Joly A, Yahiaoui I, Barthelemy D, Boujemaa N, Molino JF (2012b) The ImageCLEF 2012 Plant Identification Task. In: Forner P, Karlgren J, Womser-Hacker C, Ferro N (eds) CLEF 2012 Working Notes, CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-1178/

Goëau H, Bonnet P, Joly A, Bakic V, Barthélémy D, Boujemaa N, Molino JF (2013a) The imageclef 2013 plant identification task. In: CLEF, Valencia, Spain

Goëau H, Joly A, Bonnet P, Bakic V, Barthélémy D, Boujemaa N, Molino JF (2013b) The imageclef plant identification task 2013. In: Proceedings of the 2nd ACM international workshop on Multimedia analysis for ecological data, ACM, pp 23–28

Goëau H, Joly A, Bonnet P, Selmi S, Molino JF, Barthélémy D, Boujemaa N (2014) The lifeclef 2014 plant images identification task. In: CLEF, Sheffield, UK

Goëau H, Bonnet P, Joly A (2015) The lifeclef 2015 plant images identification task. In: CLEF, Toulouse, France

Goëau H, Bonnet P, Joly A (2016) The lifeclef plant identification task 2016. In: CEUR-WS (ed) CLEF, Evora, Portugal, CLEF2016 working notes

Goëau H, Glotin H, Vellinga W, Planqué R, Joly A (2016) Lifeclef bird identification task 2016: The arrival of deep learning. In: Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016., pp 440–449, URL http://ceur-ws.org/Vol-1609/16090440.pdf

Goëau H, Bonnet P, Joly A (2017) Plant identification based on noisy web data: the amazing performance of deep learning (lifeclef 2017). CLEF working notes 2017

Goëau H, Glotin H, Vellinga W, Planquè B, Joly A (2017) Lifeclef bird identification task 2017. In: Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017., URL http://ceur-ws.org/Vol-1866/invited_paper_8.pdf

Joalland P, Paris S, Glotin H (2014) Efficient instance-based fish species visual identification by global representation. In: Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014., pp 785–789, URL http://ceur-ws.org/Vol-1180/CLEF2014wn-Life-JoallandEt2014.pdf

Joly A, Champ J, Buisson O (2014a) Instance-based bird species identification with undiscriminant features pruning - lifeclef2014. In: Working notes of CLEF 2014 conference

Joly A, Goëau H, Bonnet P, Bakić V, Barbe J, Selmi S, Yahiaoui I, Carré J, Mouysset E, Molino JF, et al (2014b) Interactive plant identification based on social image data. Ecological Informatics 23:22–34

Joly A, Lombardo JC, Champ J, Saloma A (2016) Unsupervised individual whales identification: spot the difference in the ocean. In: Working Notes of CLEF 2016 (Cross Language Evaluation Forum)

Joly A, Goëau H, Glotin H, Spampinato C, Bonnet P, Vellinga WP, Lombardo JC, Planqué R, Palazzo S, Müller H (2017) Lifeclef 2017 lab overview: multimedia species identification challenges. In: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, pp 255–274

Kahl S, Wilhelm-Stein T, Hussein H, Klinck H, Kowerko D, Ritter M, Eibl M (2017) Large-scale bird sound classification using convolutional neural networks. In: CLEF 2017

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105

Kumar N, Belhumeur PN, Biswas A, Jacobs DW, Kress WJ, Lopez IC, Soares JVB (2012) Leafsnap: A computer vision system for automatic plant species identification. In: European Conference on Computer Vision, pp 502–516

Lasseck M (2015) Towards Automatic Large-Scale Identification of Birds in Audio Recordings. In: Mothe J, Savoy J, Kamps J, Pinel-Sauvagnat K, Jones GJF, SanJuan E, Cappellato L, Ferro N (eds) Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of

the Sixth International Conference of the CLEF Association (CLEF 2015), Lecture Notes in Computer Science (LNCS) 9283, Springer, Heidelberg, Germany, pp 364–375

Lee DJ, Schoenberger RB, Shiozawa D, Xu X, Zhan P (2004) Contour matching for a fish recognition and migration-monitoring system. In: Optics East, International Society for Optics and Photonics, pp 37–48

Nakayama H (2013) Nlab-utokyo at imageclef 2013 plant identification task. In: CLEF 2013

Nilsback ME, Zisserman A (2008) Automated flower classification over a large number of classes. In: Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing

Sánchez J, Perronnin F, Mensink T, Verbeek J (2013) Image classification with the fisher vector: Theory and practice. International journal of computer vision 105(3):222–245

Sevilla A, Glotin H (2017) Audio bird classification with inception-v4 extended with time time-frequency attention mechanisms. In: Working Notes CLEF 2017, Conf. of the Evaluation Forum, Dublin., URL http://ceur-ws.org/Vol-1866/paper_177.pdf

Silvertown J, Harvey M, Greenwood R, Dodd M, Rosewell J, Rebelo T, Ansine J, McConway K (2015) Crowdsourcing the identification of organisms: A case-study of ispot. ZooKeys (480):125

Spampinato C, Palazzo S, Joalland P, Paris S, Glotin H, Blanc K, Lingrand D, Precioso F (2016) Fine-grained object recognition in underwater visual data. Multimedia Tools Appl 75(3):1701–1720, DOI 10.1007/s11042-015-2601-x, URL https://doi.org/10.1007/s11042-015-2601-x

Stowell D, Wood M, Stylianou Y, Glotin H (2016) Bird detection in audio: A survey and a challenge. In: 26th IEEE Int. Workshop on Machine Learning for Signal Proc., MLSP, Italy, pp 1–6, URL https://doi.org/10.1109/MLSP.2016.7738875

Sullivan BL, Aycrigg JL, Barry JH, Bonney RE, Bruns N, Cooper CB, Damoulas T, Dhondt AA, Dietterich T, Farnsworth A, et al (2014) The ebird enterprise: an integrated approach to development and application of citizen science. Biological Conservation 169:31–40

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1–9

Towsey M, Planitz B, Nantes A, Wimmer J, Roe P (2012) A toolbox for animal call recognition. Bioacoustics 21(2):107–125

Trifa VM, Kirschel AN, Taylor CE, Vallejo EE (2008) Automated species recognition of antbirds in a mexican rainforest using hidden markov models. The Journal of the Acoustical Society of America 123:2424

Voorhees EM, et al (1999) The trec-8 question answering track report. In: Trec, vol 99, pp 77–82

Zhuang P, Xing L, Liu Y, Guo S, Qiao Y (2017) Marine animal detection and recognition with advanced deep learning models. Working Notes of CLEF 2017